Department of Computer Science and Engineering, National Institute of Technology Patna

End Semester Examination – Even Sem, May 2021

Answer ALL questions. This exam has three pages.

B.Tech (CS) 6th Sem

Subject: Data Mining and Warehousing **F.M:40**

Code: CS6403 **Time: 2hrs**

1. Transform the following data into a decision tree based on Gini impurity criteria

Deadline?	Is there a party?	Lazy?	Activity
Urgent	Yes	Yes	Party
Urgent	No	Yes	Study
Near	Yes	Yes	Party
None	Yes	No	Party
None	No	Yes	Pub
None	Yes	No	Party
Near	No	No	Study
Near	No	Yes	TV
Near	Yes	Yes	Party
Urgent	No	No	Study

Compute accuracy based on the above decision tree construction. If there is no deadline, there is no party happening and the person is feeling lazy, what will he do based on your constructed decision tree?

[5+1+1 marks]

2. (i) With an example of 1D data, explain how *k-means* algorithm is susceptible of getting stuck in local minima. What can be done to avoid getting stuck in local minima. **[4 marks]**
(ii) In SOM, sometimes boundaries make sense, sometimes not. Explain with examples. **[2 marks]**

3. (i) Partition the following points in 2D into two clusters using k-means algorithm.

Point↓	X-coord	Y-coord
1	0	8
2	10	1
3	9	8
4	0	10
5	10	9
6	1	4
7	0	1

Let points 2 and 5 be the initial cluster centers. Show your calculations. **[4 marks]**
(ii) In SOM, why the size of neighborhood is initially large and then it gradually becomes small as the training proceeds. **[2 marks]**

4. (i) Apply the following filters

Filter 1

1	1	1
1	1	1
1	1	1

Filter 2

0	-1	0
-1	5	-1
0	-1	0

Filter 3

1	1	1
1	-8	1
1	1	1

On cell (3,3) of the following input 2D data

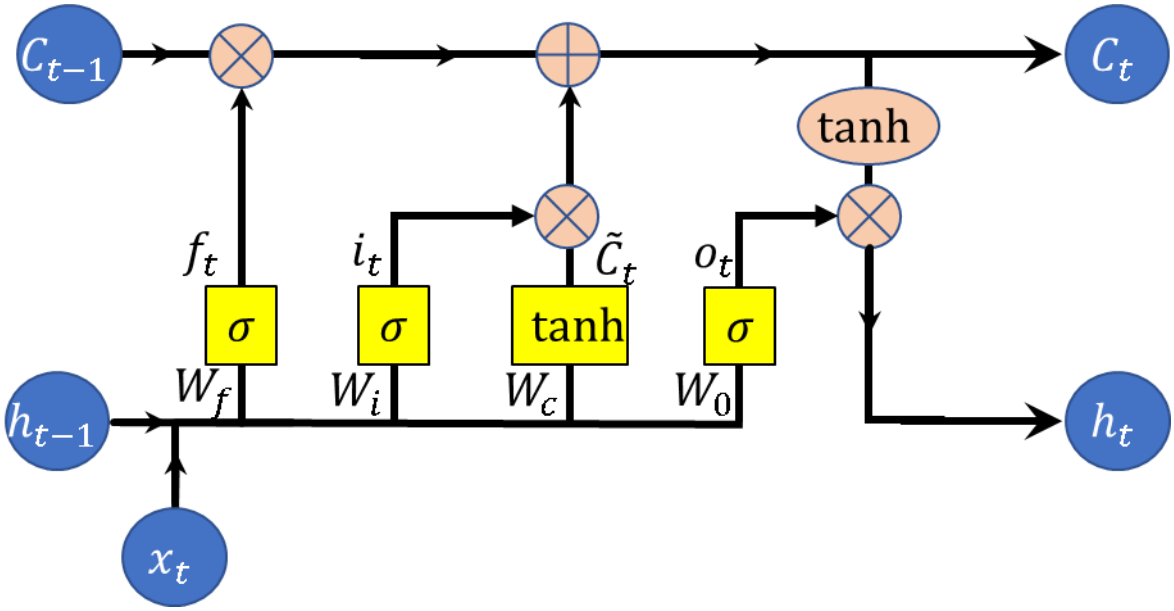
	1	2	3	4	5
1	1	0	1	1	1
2	0	0	0	1	0
3	0	0	1	0	1
4	1	1	1	1	0
5	1	1	1	1	0

Also explain what effect the above filters will have on a 2D image. **[3+1 marks]**

(ii) Give the intuition behind input gate (selective read), forget gate (selective forget), and output gate (selective write) in LSTMs. **[3 marks]**

5. (i) Why the sigmoid activation function is unable to prevent the vanishing gradient problem. Why is Relu activation function helpful towards mitigating the vanishing gradient problem. **[3 marks]**

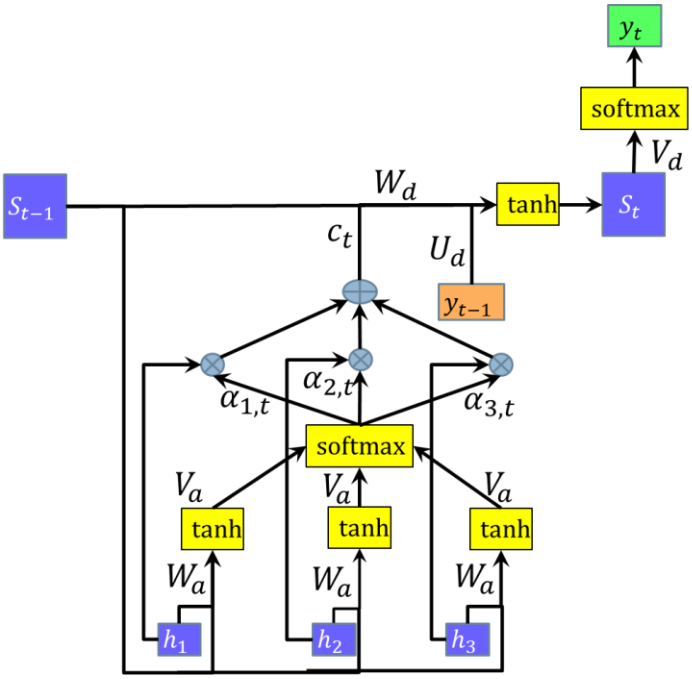
(ii) The typical LSTM network architecture is shown below:



If x_t is of dimension 3×1 and h_t is of dimension 5×1, then what is the total number of trainable parameters in the above LSTM cell. Ignore bias weights. Show your calculations. In an LSTM, is the dimension of hidden state always same as the cell state? If yes, why? **[2+1 marks]**

6. (i) Motivate the need of and explain the attention mechanism with an example. **[4 marks]**

(il) Consider the following attention mechanism proposed by Bahdanau et. al. 2015.



$$e_{jt} = V_a \cdot \tanh(W_a(S_{t-1}, h_j))$$

$$\alpha_{jt} = \text{softmax}(e_{jt})$$

$$c_t = \sum_{j=1}^{T} \alpha_{jt} h_j$$

$$S_t = \tanh(W_d(S_{t-1}, c_t) + U_d y_{t-1}))$$

$$y_t = \text{softmax}(V_d S_t)$$

Given that each of h_1, h_2, h_3, and S_t is of size 3×1, V_a is of size 2×1, and y_t is of size 5×1, compute the total number of trainable parameters in the above attention mechanism. Ignore bias weights. Show your calculations. **[3 marks]**

(iii) In RNNs the notion of order in sequential data is preserved by the way RNNs process data in a time-based sequential manner. However, transformer network takes the entire input at once and does not have any recurrent connections. How does the transformer network capture the sequential information then? **[1 marks]**